Fake Product Review Monitoring And Removal

Eknath Eswar Reddy Bhavanam
Department of Computer Science,
University of Bridgeport, Bridgeport, CT

Introduction

Today, the most common habit that people these days have is to read online opinions/reviews for different purposes like before going for shopping, reading books, renting a car, going on a trip, etc. Today many people read/write reviews on merchant sites, blogs, forums, and social media before/after they purchase products or services. If you want to buy a product, you would typically go to a review site like "amazon.com" to read some reviews of the product you would like to buy. When you find too many positive reviews, you are more likely to buy the product and when you find more negative reviews about the product you will certainly not buy it. Positive reviews also result in huge financial gains & fames for businesses, companies, organizations and individuals, which, unfortunately, gives strong incentives for frauds to display fake reviews to promote or to discredit some target products or services. While, negative reviews can damage reputation and cause monetary loss. Such frauds are called opinion spammers and their activities are called opinion spamming. These spam reviews come in two forms: defaming-spam which untruthfully vilifies, or hype spam that deceitfully promotes the target product.

Problem Definition

Simply put, we consider the problem of spotting fraudulent reviewers, and consequently spotting fake reviews in online review datasets. The online review datasets mainly consist of a set of users (also called customers, reviewers), a set of products (e.g., hotels, restaurants, etc.), and the reviews. Each review is written from a particular user to a particular product, and contains a star-rating, often an integer from 1 to 5. As such, a review dataset can be represented as a bipartite network. In this network, user nodes are connected to product nodes, in which the links represent the "reviewed" relationships and each link is associated with the review rating. The objects in the review network, i.e. the users, products, and reviews, can be grouped into certain classes. In this paper, we consider two classes for each object type: products are either good or bad quality, users are either honest or fraud, and finally reviews are either real or fake. Intuitively, a product is good (bad) if it most of the time receives many positive (negative) reviews from honest users. Similarly, a user is honest (fraud) if s/he mostly writes positive (negative) reviews to good products, and negative (positive) reviews to bad products. In other words, a user is fraud if s/he is trying to promote a set of target bad products (hype spam), and/or damage the reputation of a set of target good products (defaming-spam). All the reviews of honest users can safely be regarded as real. In an ideal setting, all the reviews of the fraud users can also be thought of as fake. However, in reality fraudsters could also write realistic reviews, trying to hide their otherwise fraudulent activity. All in all, notice that the class labels of the interconnected objects in the review data are strongly correlated as they are described in terms of one another. As a result, it is natural to think of a review dataset as a network in order to attack the opinion fraud problem coherently.

Dataset and Data Analysis

The Amazon product reviews data set collected by crawling all the software product (app) reviews under the Amazon Fine Food category from www.kraggle.com. This data was originally published on SNAP. The Amazon Fine Food Reviews dataset consists of 568,454 food reviews Amazon users left up to October 2012. As part of a review, a user rates a product from 1 (worst) to 5 (best) and the few lines of comments. The dataset consists of data related to product id, user id, score, comment summary.
We use the data set from Amazon product review data.

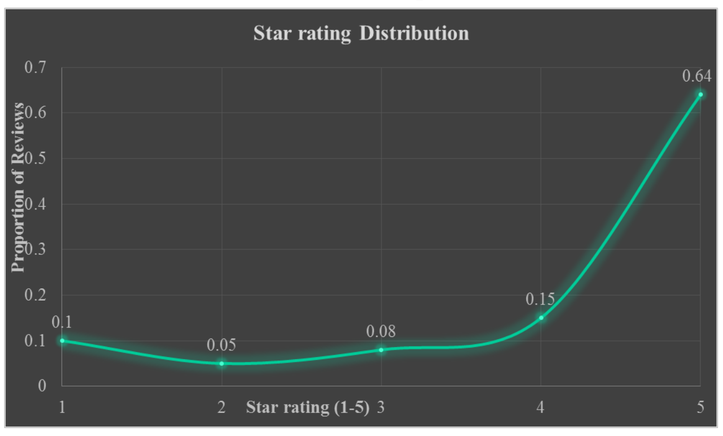

Figure 1 : Star rating Distribution of the Amazon Fine Food Reviews dataset. As expected, the distribution has a characteristic 'J' shape, which reflects user apathy when deciding whether to review products. In this dataset, the reviews are skewed towards positive ratings.

Synthetic Dataset

In order to validate the effectiveness of our algorithm, we first tested it on a synthetically generated review dataset.
In the synthetic dataset we used there are 4 products namely p1, p2, p3 and p4 where products p2 and p4 are bad products and p1 and p3 are good products. And 3 users with user id u1, u2 and u3 and assigned u2 are fraudster and remaining honest users. We also tried to put the same J shape for the star rating distribution similar to the original dataset.

Id	Product Id	User Id	Rating
1	p1	u1	5
2	p2	u1	3
3	p1	u2	1
4	p3	u1	5
5	p2	u2	5
6	p1	u3	5
7	p3	u3	4
8	p4	u2	5
9	p4	u3	3
10	p2	u3	2

Table 1 : shows the synthetic dataset

Problem definition

We can define Fake product review Monitoring and removal problem more formally.
Given,
- a bipartite network $G_s = (V, E)$ of users and products connected with signed edges,
- prior knowledge (probabilities) of network objects belonging to each class, and
- compatibility of two objects with a given pair of labels being connected;

Classify,
The network objects $Y_i \in Y = Y^V \cup Y^E$, into one of two respective classes; $L_U = \{honest, fraud\}$, $L_P = \{good, bad\}$, and $L_E = \{real, fake\}$, where the assignments y_i maximize the objective probability in the Equation (1) below

$$p(y|x) = \frac{1}{Z(x)} \prod_{Y_i \in Y_v} \Psi_i\ (y_i) \prod_{e(Y_i, Y_v, s) \in \varepsilon} \Psi_{ij}^{s}(y_i, j) \quad (1)$$

We propose a new algorithm that extends Loopy Belief Propagation (LBP) in order to handle signed networks. At convergence, we use the maximum likelihood label probabilities for scoring.
signed Inference Algorithm (sIA) The inference algorithm applied on a signed bipartite network can be concisely expressed as the following equations:

$$m_{i \to j}(y_j) = \alpha 1 \sum_{y i \in \mathcal{L} u} \Psi_{ij}^{s}(y_i, y_j) \Psi_i^u(y_i) \prod_{Y_k \in N_i \cap y^p \setminus Y_j} m_{k \to i}(y_i), \forall y_j \in \mathcal{L}^p \quad (2)$$

$$b_i(y_i) = \alpha_2 \Psi_i^u(y_i) \prod_{Y_k \in N_i \cap y^p} m_{j \to i}(y_i), \forall y_j \in \mathcal{L}^u \quad (3)$$

where $m_{i \to j}$ is a message sent by user Y^i to product Y^j (a similar equation can be written for messages from products to users), and $b_i(y_i)$ denotes the belief of user i having label y_i (again, a similar equation can be written for beliefs of products). α's are the normalization constants, which respectively ensure that each message and each set of marginal probabilities sum to 1.

Evaluation Plan

We first use the proposed algorithm on the synthetic dataset The sentiment on edges are then assigned as follows. If there is an edge in the synthetic graph, i) honest users always give '-' to bad products, ii) fraudsters always give '+' to bad products, iii) fraudsters always give '+' to the famous good products (to hide their otherwise bad activity), and iv) honest users always give '+' to good products. This way we will be getting 3 fake reviews.
Proposed algorithm result. We show the class member ships for top-scoring (most malicious) users and products found by our sIA in top row of Figure 2. In (a), the algorithm successfully ranks all the fraudsters on top. In (b), all bad products are also ranked top with very high scores, while another product also shows up with high score (0.75) —this product has degree 1, from a fraudster with a '+' review, which increases its probability of being bad. Results are similar for fake reviews, which we omit for brevity.

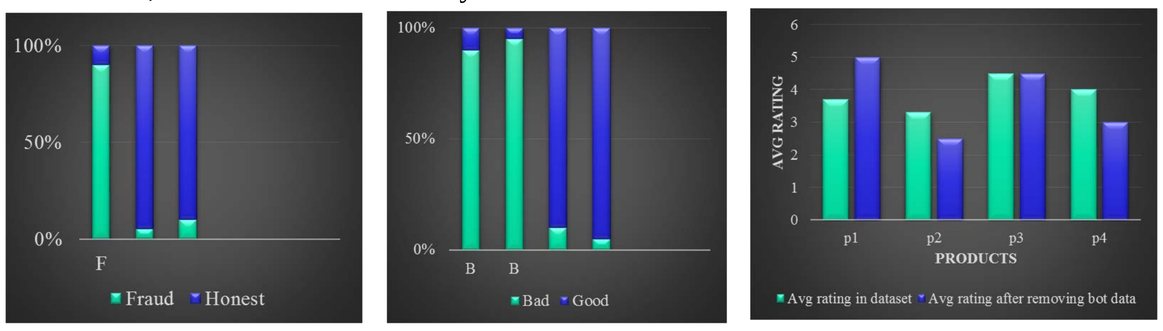

Figure 2 : (a) (b) (c)

The fraudsters in the detected bot, all with rating 5, significantly affect the average rating of the 5 products they reviewed. In Figure 2 (c), notice that all their average ratings changes, once those fraudsters and their reviews are removed from our dataset.

Conclusion

We propose an algorithm that exploits the network effects to automatically detect fraudulent users and fake reviews in online review networks. Main contribution are:
Problem formulation: We formally define the Fake product review monitoring and removal problem as a classification task on signed bipartite review networks, and thus we capture the network effects for improved classification.
Scoring algorithm: We show how to efficiently solve the inference problem, on signed bipartite networks. Our approach uses several compatibility matrices and computes scores for all 3 types of objects: reviews (fake/truthful), users (honest/fraud), and products (good/bad quality).
Evaluation: We worked our algorithm on synthetic as well as real networks. This algorithm successfully detects fraudulent attack groups, and the users that significantly distort product ratings.

References

Opinion Fraud Detection in Online Reviews by Network Effects
Leman Akoglu, Rishi Chandy, Christos Faloutsos, 2013
Inferring networks of substitutable and complementary products
J. McAuley, R. Pandey, J. Leskovec
Knowledge Discovery and Data Mining, 2015
Image-based recommendations on styles and substitutes
J. McAuley, C. Targett, J. Shi, A. van den Hengel
SIGIR, 2015
Jindal and Liu, WWW-2007; WSDM-2008; Lim et al, CIKM-2010; Jindal, Liu and Lim, CIKM-2010; Mukherjee et al. WWW-2011; Mukherjee, Liu and Glance, WWW-2012